

R für Dummies - Alles-in-einem-Band

Grundlagen und fortgeschrittene Programmier Techniken

» Hier geht's
direkt
zum Buch

DAS VORWORT

Einführung

In diesem Buch habe ich alle Informationen zusammengetragen, die Sie benötigen, um mit R auf Anhieb zurechtzukommen. Der Schwerpunkt liegt natürlich auf dem Thema Statistik, da die Entwickler von R diese Sprache für die Datenanalyse entwickelt haben.

Sie sollten also die Grundlagen der Statistik verstehen. Und gleich vorab: Dieses *Alles-in-einem-Band-Buch* beschreibt keine vorgefertigten Lösungen oder Antworten. Ich habe Statistik nie auf diese Weise gelehrt und werde es auch nie so tun. Bevor ich Ihnen zeige, wie Sie R einsetzen, um mit einem statistischen Konzept zu arbeiten, vermittele ich Ihnen eine solide Grundlage, worum es bei diesem Konzept überhaupt geht.

Tatsächlich sind die Teile 2 und 3 dieses Kompendiums eine Einführung in die Statistik, wobei zufällig R als Mittel zur Erklärung statistischer Ideen verwendet wird.

Teil 4 wendet ebenfalls diese Methode an. Dort werden die Ideen hinter dem maschinellen Lernen vermittelt, bevor Sie erfahren, wie R verwendet wird, um sie zu implementieren. Teil 5 präsentiert eine Reihe von Projekten, anhand derer Sie Ihre neu erworbenen R-Kenntnisse praxisnah üben können.

Möchten Sie mehr erfahren? Lesen Sie weiter!

Über dieses Buch

Dieses Buch (und gegebenenfalls das E-Book) besteht aus fünf Teilen, die mehrere Dimensionen von R abdecken.

Teil 1: Einführung in R

Wie bereits erwähnt, beschäftigt sich die Sprache R mit Statistik. Dementsprechend bietet Teil 1 eine Einführung in die grundlegenden Konzepte der Statistik, die Sie einfach kennen *müssen*, um mit R voranzukommen.

Anschließend lernen Sie R und RStudio kennen. RStudio ist eine weit verbreitete Entwicklungsumgebung für die Arbeit mit R. Zunächst beschreibe ich die Grundlagen des R-Codes und gehe auf R-Funktionen und -Strukturen ein.

Richtig lebendig wird R, wenn Sie die speziellen Packages verwenden, die Sie schon früh kennenlernen.

Teil 2: Daten beschreiben

Bei der Arbeit mit Statistiken geht es oft darum, Daten auf sinnvolle Weise zusammenzufassen. In Teil 2 erfahren Sie, wie Sie dazu vorgehen.

Die meisten Menschen wissen, was Durchschnittswerte sind und wie man sie berechnet. Aber das ist nicht die ganze Geschichte. In Teil 2 erzähle ich Ihnen von zusätzlichen deskriptiven Statistiken, die die Lücken füllen, und ich zeige Ihnen, wie Sie R verwenden, um diese Statistiken zu berechnen und damit zu arbeiten. Außerdem lernen Sie, Grafiken zu erstellen, die die Datenbeschreibungen und -analysen visualisieren, die Sie in Teil 2 und 3 kennenlernen.

Teil 3: Daten auswerten

Teil 3 befasst sich mit dem grundlegenden Ziel der statistischen Analyse: über die reinen Daten hinauszugehen und Ihnen dabei zu helfen, Entscheidungen zu treffen. In der Regel handelt es sich bei den Daten um Messungen einer Stichprobe, die einer großen Grundgesamtheit entnommen wurde. Das Ziel ist es, anhand dieser Daten herauszufinden, was in der Grundgesamtheit vor sich geht.

Dies wirft eine ganze Reihe von Fragen auf: Was bedeutet ein Durchschnitt? Was bedeutet die Differenz zwischen zwei Durchschnittswerten? Sind zwei Dinge miteinander verbunden? Dies sind nur einige der Fragen, die ich in Teil 3 anspreche. Hier werden Sie auch lernen, die R-Tools anzuwenden, die Ihnen helfen, diese Fragen zu beantworten.

Teil 4: Aus Daten lernen

Die Erstellung effektiver Modelle für maschinelles Lernen erfordert Erfahrung. In Teil 4 sammeln Sie Erfahrung anhand von Projekten zum maschinellen Lernen. Zusätzlich zu den Projekten, die Sie zusammen mit mir durchführen, schlage ich Ihnen weitere Projekte vor, an denen Sie sich ausprobieren können.

Zu Beginn werde ich Ihnen das University of California-Irvine Machine Learning Repository vorstellen, das die Datensätze für die meisten Projekte in Teil 4 bereitstellt.

Um Ihnen einen sanften Einstieg zu ermöglichen, zeige ich Ihnen das Rattle-Package zur Erstellung von Anwendungen für maschinelles Lernen. Es ist eine freundliche Schnittstelle zu den Funktionen von R für maschinelles Lernen. Ich mag `Rattle` sehr, und ich gehe davon aus, Sie werden es auch mögen. Sie verwenden es, um Entscheidungsbäume, Zufallswälder, Support-Vektor-Maschinen, den k-Means-Algorithmus und neuronale Netze kennenzulernen und mit ihnen zu arbeiten.

Sie werden mit sehr großen Datensätzen arbeiten – nicht die Terabytes und Petabytes, mit denen es Datenwissenschaftler zu tun haben, aber groß genug, um den Einstieg zu finden. In einem Projekt analysieren Sie einen Datensatz von mehr als 500 000 Flügen einer Fluggesellschaft. In einem anderen Projekt führen Sie eine Kundensegmentierungsanalyse von über 300 000 Kunden eines Online-Händlers durch.

Teil 5: R nutzen – Einige Projekte, die Sie auf die Praxis vorbereiten

Wie der Titel vermuten lässt, ist auch Teil 5 projektorientiert aufgebaut.

In den hier vorgestellten Projekten erstellen Sie Anwendungen, die auf Benutzer reagieren. Ich zeige Ihnen das Package `shiny` für die Arbeit mit Webbrowsern und das Package `shinydashboard` für die Erstellung von Dashboards.

All dies führt Sie weg von dem ursprünglichen Lebensauftrag von R, aber Sie bekommen eine Vorstellung davon, wie Sie sich mit R in völlig neue Richtungen entwickeln können.

Wenn Sie eine Weile mit R gearbeitet haben, werden Sie vielleicht einige dieser neuen Richtungen für sich entdecken!

Was Sie getrost überspringen können

Jedes Nachschlagewerk bietet eine Fülle von Informationen; dieses Buch bildet keine Ausnahme. Ich wollte, dass alle bereitgestellten Informationen nützlich sind, aber ich habe nicht alles auf das gleiche Niveau ausgerichtet. Wenn Sie also nicht tief in die Materie eindringen wollen, können Sie die mit dem Symbol für technisches Material gekennzeichneten Abschnitte überspringen und auch die Einschübe auslassen.

Törichte Annahmen über den Leser

Ich gehe davon aus, dass

- ✓ Sie wissen, wie man mit Windows oder dem Mac arbeitet. Ich gehe nicht auf die Details von Zeigen, Klicken, Auswählen usw. ein.
- ✓ Sie R und RStudio installieren (wie das geht, erkläre ich in Teil 1) und den Beispielen folgen können. Ich verwende die Windows-Version von RStudio, aber Sie sollten keine Probleme haben, wenn Sie mit einem Mac arbeiten.

In diesem Buch verwendete Symbole

Wie in allen *Für-Dummies*-Büchern helfen Ihnen Symbole bei Ihrer Reise. Sie werden am Seitenrand angezeigt, um Sie auf bestimmte Absätze aufmerksam zu machen.



Dieses Symbol weist auf einen Hinweis oder eine Abkürzung hin, die Ihnen bei Ihrer Arbeit helfen können.



Dieses Buch zeigt zeitlose Weisheiten auf, die Sie auf Ihrem Weg zur Erleuchtung mitnehmen können.



Achten Sie auf dieses Symbol! Es macht Sie darauf aufmerksam, etwas zu vermeiden, das Ihnen die Arbeit erschweren könnte.



Wie ich bereits in »Was Sie getrost überspringen können« erwähnt habe, zeigt dieses Symbol an, dass Sie den Inhalt überblättern können, wenn er zu technisch ist. (Ich habe diese Inhalte auf ein Minimum beschränkt.)

Wie es jetzt weitergeht

Legen Sie los! Sie können überall anfangen, aber hier ein paar Tipps. Sie möchten sich in R und Packages einarbeiten? Dann ist Teil 1 genau das Richtige für Sie. Ihre letzte Statistikvorlesung ist schon eine Weile her (oder vielleicht haben Sie noch nie eine besucht)? Dann fangen Sie in Teil 2 an. Für alles andere gucken Sie im Inhaltsverzeichnis oder im Index nach und greifen gezielt auf die benötigten Informationen zu.

Wenn Sie lieber von der ersten bis zur letzten Seite lesen möchten, blättern Sie einfach weiter ...